Communications of the Association for Information Systems

Volume 2 Article 16

September 1999

Developing and Delivering a Data Warehousing and Mining Course

Elizabeth M. Pierce

Indiana University of Pennsylvania, empierce@grove.iup.edu


Communications of the Association for Information Systems

Volume 2, Article 16
September 1999

DEVELOPING AND DELIVERING A DATA WAREHOUSING AND MINING COURSE

Elizabeth M. Pierce
MIS & Decision Sciences Department
Indiana University of Pennsylvania
empierce@grove.iup.edu

TEACHING

DEVELOPING AND DELIVERING A DATA WAREHOUSING AND MINING COURSE

Elizabeth M. Pierce
MIS & Decision Sciences Department
Indiana University of Pennsylvania
empierce@grove.iup.edu

ABSTRACT

This paper describes the content and delivery of a Data Warehousing and Mining course that was developed for students in the Eberly College of Business at Indiana University of Pennsylvania. This elective course introduces students to the strategies, technologies, and techniques associated with this growing MIS specialty area. Students learn what is involved in planning, designing, building, using, and managing a data warehouse. Students also learn about how a data warehouse must fit into an over-all corporate data architecture that may include legacy systems, operational data stores, enterprise data warehouses, and data marts. In addition, students are exposed to the different data mining techniques used by organizations to derive information from the data warehouse for strategic and long-term business decision making.

Keywords: IS curriculum, database, data warehouse, data mining

I. INTRODUCTION

Over the last several years, interest in data warehouses and data mining techniques increased markedly. A quick survey of ComputerWorld's News and

Features archives (www.computerworld.com) shows just how much that interest grew over the last five years.

Table 1. Result of Search of ComputerWorld Archive's

Year	Search Phrase	Number of Hits	Search Phrase	Number of Hits
1994	Data warehouse	47	Data mining	5
1995	Data warehouse	157	Data mining	27
1996	Data warehouse	259	Data mining	79
1997	Data warehouse	409	Data mining	90
1998	Data warehouse	359	Data mining	86
1/1/99 - 9/15/99	Data warehouse	194	Data mining	53

Data warehouses and data mining are not new concepts. A data warehouse is a large database where data from many different sources (e.g., billing records, scanner data, registration forms, call records, coupon redemptions, surveys) are gathered together and organized into a consistent format. Data mining refers to a family of quantitative techniques that individuals can use to analyze data for use in business decision making. Yet, what some may have considered as just another database application is now a whole new methodology for how companies should organize and use their data.

Data mining gained its current popularity as a result of its many interesting applications. Companies today successfully apply data mining techniques to detect fraud, to study customer churn, to weed out bad customers, to customize marketing efforts, to approve new customers, to improve inventory control, to develop new products and services, and to strengthen customer relationships. For companies to use data mining to gain competitive advantage, however, they must have three ingredients in place:

(1) computing power,

(2) commercial data mining software packages, and

(3) access to large volumes of good quality data.

Improvements in computers and new software packages satisfy requirements 1 and 2. The development of the data warehouse satisfies the third criterion.

II. COURSE HISTORY AT IUP

As the literature on data warehousing and mining grew, it seemed appropriate to develop a course to teach students about this growing MIS area. "Data Warehousing & Mining" was first taught as a special topic in the Fall 1998 semester at the Eberly College of Business at Indiana University of Pennsylvania (IUP). It was offered a second time in Spring 1998 and will be offered as an elective at least once a year in the future. As a special topic, the course is open to both upper level undergraduate students in MIS and MBA students with an interest in information technology.

At the Eberly College of Business, the major requirements for undergraduate MIS students are based on the IS '97 curriculum guidelines. Our MIS curriculum consists of programming languages such as Visual Basic and Cobol, computer hardware, database, systems analysis, systems design, and networking essentials. Students also are required to take at least 3 electives related to their major area. At the graduate level, MBA students, in addition to exposure to databases through undergraduate IS courses and work experiences, take a graduate level course in Information Systems which includes instruction in database theory and applications.

While students at both the undergraduate and graduate levels learn about database theory and applications as part of their curriculum requirements, we believe that data warehousing involves such a significant amount of new theory and concepts as to warrant a course in its own right. Whereas traditional database courses are designed to give students an understanding of database architecture, concepts, and features so that they can properly design databases to support operational systems, data warehousing databases have a different focus. In data warehousing, the idea is to create analytical data repositories to support strategic, managerial decision-making. In addition, data warehousing

incorporates the idea that both operational and analytical databases should be part of an over-all integrated corporate data architecture. Thus, this elective is intended for students interested in pursuing further applications in databases as well as for those students interested in understanding how to set up an over-all data strategy for an organization.

III. COURSE HISTORY AT OTHER SCHOOLS

On the Internet, many companies and consulting services prominently advertise their training programs in the different aspects of data warehousing and mining. Tracking down colleges and universities that offer these courses is more difficult. Many colleges and universities may, in fact, be offering a data warehouse and mining course, but this information may be hidden behind a special topic or seminar course number. From the successful hits that were obtained, it appears that colleges and universities employ several different strategies for exposing students to this topic.

1. Some schools include data warehousing as part of the course material for an existing DBMS, MIS, or Systems Design course. Under this strategy, the time allotted to data warehousing is short, usually a week or less. This approach enables students to gain an awareness of the topic, but does not permit in-depth coverage.

2. Some schools include data warehousing as part of an advanced data management course. In addition to data warehousing, the advanced data management course may also include coverage of distributed databases, object oriented databases, web-based databases, decision support systems, and data mining. Similar to the course at IUP, these schools require that students have prior exposure to databases or decision support systems before taking this course. A sample of schools offering this type of data warehousing course is included as Appendix A.

3. At least one school now offers a certificate program in data warehousing. UC Berkeley Extension through its adult continuing education

program (http://amber.berkeley.edu:4243/cert/data.html) offers a 40 lecture weeks (225 hours of instruction) Data Warehousing Certificate program composed of 9 different courses.

4. Because some universities use data warehouses as part of their internal decision support systems, these schools often provide instructional materials on how to query their data warehouse. While this training is not part of the academic curriculum, it does enable students and faculty to gain hands-on experience with accessing an actual data warehouse. Examples of universities who now employ data warehouses include Syracuse University, University of Michigan, IUP, and Arizona State University.

IV. COURSE OBJECTIVES

The objectives of the data warehousing course offered at IUP are:

1. By the end of the course, students should be able to describe what is meant by a data architecture and be able to explain how legacy systems, data warehouses, data marts, and operational data stores fit into that data architecture. Students need to understand that databases do not operate as separate islands of information, but rather need to be coordinated into a cohesive plan for supporting the operational, tactical, and strategic needs of the organization.

2. Students should understand the process by which data warehouses are designed, built and managed. Students also need to understand the many issues that complicate data warehouse construction and maintenance. For example, performance is a problem as data warehouses grow beyond gigabytes to terabyte, petabyte or even exabyte size. How to manage metadata so business analysts can use the stored data more effectively is also an issue. Figuring out which source of data to use and then how best to integrate and map the data is another challenge in creating a successful data warehouse. By the end of the course, students should be able to explain the steps they would follow

to create a data warehouse as well as the different types of problems and issues that they will need to address.

3. Students should be aware of the variety of end user tools for analyzing the data contained in the data warehouse. These tools range from simple spreadsheets to sophisticated neural net software packages. Students should understand the selection process for matching the right set of tools to the various types of users in an organization. Students need to understand that a data warehouse must be flexible enough to accommodate people whose abilities may range from novice to power user.

V. COURSE PREREQUISITES

The prerequisites for this course are:

- a good understanding of files and databases.
- a good understanding of business needs and practices
- a background in some type of quantitative business course such as business statistics or quantitative methods for managers.

Our department requires students to have either taken a prior course in database or through other experiences been exposed to data modeling and database concepts. Students should have already mastered how to define a database, how to manipulate data in a database, how to use the system catalog, how to use views, and how to write embedded SQL before taking this course. In addition, if students understand the internal level of the DBMS architecture then they are in a much better position to appreciate why a design for a database whose principal activity is updating is not optimal for a database whose principal activity is querying.

Data warehouses are constructed to house data from multiple business functions to support the decision-making needs of managers. Students who have the technical skills, but lack either business core courses or work experiences often have difficulties in identifying with the problems and issues that managers face when trying to coordinate diverse data into information that can be used to

help solve a business problem. For this reason, we restrict this class to senior undergraduate and MBA students.

Some type of quantitative business course, such as business statistics or quantitative methods for managers, is a prerequisite for two reasons. First, since this course involves an overview of data mining techniques, students are better prepared to learn this material if they have finished their quantitative curriculum requirements. Second, data warehouses should be designed to support quantitative analysis. Students with experience in statistical software packages can better identify with the user's data requirements for the data warehouse.

VI. WEEKLY COURSE CONTENT

We designed the content of the course to be delivered over 14 weeks assuming 3 hours of class time per week. The course begins with the evolution of data warehousing and ends with a survey of some of the better-known techniques in data mining. This schedule, shown in Table 2, can be condensed or expanded as needed to fit a different semester schedule.

To complement the lecture topics, students are assigned weekly activities that included readings, discussion questions, problems, and short papers. These activities are summarized in Table 3.

Table 2. Summary of Course Content

Week	Lecture/Discussion Topics
1	Review Database Terminology. Define and discuss the difference between operational (i.e. the data used to run the day to day business) and analytical data (data used to analyze a particular problem or situation). Discuss the evolution of the data warehouse. Explain the need for two types of databases: one designed for storing operational data, the other designed for storing analytical data.
2	Explain basic data warehouse definitions and terminology. Explain the different variations of the data warehouse including data marts, operational data stores, and enterprise data warehouses. Discuss how to choose an appropriate data warehouse architecture and infrastructure. Discuss planning issues concerning the construction of a data warehouse. List the reasons for creating a data warehouse, the criteria for success, the different roles/individuals/expertise required on the data warehouse team. Give a preview of the different phases in data warehouse construction. Discuss the various methods for gathering requirements, how to conduct a pilot study, how to determine what the subject area of the data warehouse will be, how to identify data sources, how to conduct user interviews, and how to use data modeling tools to develop the data warehouse design.
3	Discuss issues concerning data quality such as the problem of dirty data, the problem of undocumented business rules and nuances of meanings, the problem of synonyms, homonyms, analogs, and the perils of using external data. Discuss basic design concepts in data warehouses. Topics include schemas, entities, attributes, relationships, granularity, data hierarchies, and time components. Discuss whether the data warehouse will hold status data (snapshots) or event data (transactions) or both. Also discuss if the data warehouse will be centralized or decentralized. Discuss the different table designs that range from one large table, to star designs (central fact table surrounded by dimensional tables), constellation design, snowflake design, and blizzard design.
4	Discuss additional design features for data warehouses. These include incorporating value chains and conforming dimensions, handling heterogeneous entities, resolving slowly changing attributes, and using mini-dimensional tables, factless fact tables, and degenerate keys.
5	Profile several data warehouse examples in various business functional areas and industries.
6	Discuss the issues involved in cleaning and transforming data, validating and testing data, and how best to populate and update the data in the data warehouse on an automatic basis. Discuss the importance of metadata and how to use and maintain metadata in the data warehouse.

Table 2. Summary of Course Content (continued)

7	Discuss issues involved with data access and use of the data warehouse. Explore the different types of end users and how best to train users to get the most out of the data warehouse. Also discuss the wide variety of end user tools available (EIS, DSS, Report Writers, Ad Hoc Query, Application Development, Spreadsheet Analysis, Statistical Analysis, Visualization Tools, Data Mining, OLAP). Discuss the steps that the data warehouse team should follow to ensure that the company selects the right set of end user tools. Discuss data warehouse maintenance and management issues. These include how to assist the end user, how to market the data warehouse to internal customers, how to monitor the quality and performance of the data warehouse, how to charge for use of the data warehouse, how to ensure security, and how to implement backup and recovery procedures.
8	Discuss techniques for handling performance problems with the data warehouse. These techniques include summarization, archiving, partitioning, indexes, de-normalization, sampling, and parallel processing Discuss what is meant by OLAP - Online Analytical Processing and how it can be implemented either using a relational database model or using a multidimensional database model.
9	Discuss the development of data mining. Discuss the steps one follows in setting up a data mining study and provide an overview of the different techniques and tools that can be employed. Explain how market basket analysis works.
10	Explain how memory based reasoning and cluster techniques work.
11	Explain how link and decision tree analysis work.
12	Explain how neural nets work.
13	Explain how genetic algorithms and visualization techniques work.
14	Discuss future trends and topics in data warehousing such as data warehousing and the Internet and handling objects (video, audio, graphics, images) in the data warehouse.

Table 3. Summary of Course Activities

Activity	Description
Readings	In addition to their textbooks, students can read articles from trade journals or the Internet. The following sites are useful sources for information on data warehousing and mining. • **www.brint.com** - This site contains white papers on a number of MIS topics including data warehousing and data mining. • **www.computerworld.com** - This is a good source for articles on MIS topics including data warehousing and data mining. • **www.data-miners.com** - This web site is good place to start when looking for information on data mining tools. • **www.datawarehouse.com** - This web site promotes itself as one-stop resource for industry professionals interested in keeping abreast on news of data warehousing and data mining. • **www.datawarehouse.org** - This web site is an online source for information about data warehouse technologies. • **www.dbaint.com** - Through their web site, DataBase Associates provides access to articles and papers containing technical and product information in three key areas of information technology: data warehousing, business intelligence tools and analytic applications, and Web-based information distribution. • **www.DMReview.com** - This web site contains the archives of the Data Management Review. This magazine covers the data warehousing field and includes leading figures on data warehousing among its columnists. • **www.dw-institute.com** - The Data Warehousing Institute's web site contains free white papers that anyone can request. • **www.kdnuggets.com/subscribe.html** - The subscribe.html allows receiving a weekly listserve. Remove the subscribe.html and students reach a major data source for data mining. • **pwp.starnetinc.com/larryg** - This is the URL for the Data Warehousing Information Center run by Larry Greenberg. • **power.cba.uni.edu/dss/dwolap.html** - This site is part of a general site on decision support systems maintained by Dan Power at the University of Northern Iowa. It provides access to other sites and has a list of white papers, particularly on OLAP applications. • **www.sas.com/software/data_warehouse** offers news, events, white papers, demos, documentation, training, and consulting on data warehousing. • **www.sas.com/software/data_mining** offers news, events, white papers, training, and consulting on data mining. • **www.spss.com/datamine** - This web site offers software demonstration versions, movies, white papers, and data interfaces on data mining.
Discussion Questions	Students are assigned review and discussion questions on a weekly basis to test their understanding of the concepts.

Table 3. Summary of Course Activities (continued)

Projects	Although students are anxious to develop skills in the major DBMS (Oracle, Sybase, Informix, SQL Server) that are used in data warehouse construction, the lack of a sophisticated DBMS is not fatal to the course. In our labs, students used Microsoft Access to develop their own mini-data warehouses as a way to practice the design principles for constructing a data warehouse.
Papers	Because this field is continuing to evolve, it is important that students learn how to seek out information on their own. During the course of the semester, students are responsible for conducting a literature search for a particular topic and then writing up of their findings. Sometimes the topics were assigned. Sometimes students picked their own data warehouse topic.
Speakers	At IUP there is a data warehouse project under way so that administrators can better understand enrollment patterns. This project provided an excellent source for speakers to talk to the class about their experiences. Local companies are another good source for speakers.

VII. TEXTBOOK SUPPORT

The first books about data warehousing appeared in the early 1990's. William Inmon [1996] who is often credited as being "the father of the data warehouse" was one of the first to define and explain this new concept in storing, managing, and using data. He and his co-authors wrote a series of books that cover in detail the evolution of the data warehouse from its early conception to today's interpretation.

One of the challenges in teaching a data warehouse course at the present time is that while many books are now available on the topic, the vast majority are trade books written for the IS professional rather than the IS student. This means that the books often assume a familiarity with corporate business practices that many college students lack. These books also lack the exercises, discussion questions, cases, and other amenities usually found in academic textbooks to help guide the student (and the instructor) through the material. Finally, while many trade books provide an excellent conceptual picture of data warehouses, many lack the specifics that students are hoping to acquire as to how one actually sets up a data warehouse.

For the course, the instructor chose two trade books that she felt came closest to the ideal textbook. Kimball's *The Data Warehouse Toolkit* [1996]

provides a good overview of the steps one must go through to design, build, and implement a data warehouse. His book also had the nice feature of including a CD-ROM with working models for all the databases described in the book as well as software for querying the dimensional data warehouses.

The second book chosen for the course is *Data Mining Techniques* by Berry and Linoff [1997]. This book gives the students a good overview of the techniques used to uncover patterns in data to improve business decision making. Because data mining is intertwined with the concept of data warehousing, our course seeks to cover both areas. It is quite possible as this subject matter evolves that two separate courses may be needed in the future to provide adequate coverage of the subject matter. In practice, most trade books either deal with the one issue (data warehouse construction) or the other (what can one do with the data in the data warehouse), hence the need for two books for the course.

While these two books provide the bulk of the material for the course, they lack depth in some of the organizational and management issues surrounding data warehouses. To remedy these gaps, the instructor used material supplied by several other authors in the course. Two excellent books are worth noting. *Data Warehouse: Practical Advice from the Experts* [Bischoff and Alexander,1997] is a compilation of practical advice and issues to consider when building a data warehouse. *Building a Data Warehouse for Decision Support* [Poe, Klauer, and Brobst, 1998] provides a good understanding of how data warehouses must fit into an over-all data architecture for supporting the decision making needs of the organization. A complete list of the books used to develop the course is in the Bibliography section of this paper.

VIII. CHALLENGES IN COURSE DESIGN AND DELIVERY

An instructor will face some additional challenges besides the lack of data warehousing and mining textbooks and supplemental course materials when trying to teach this course. One issue involves the changing subject matter.

Because this area is still relatively new, the concepts continue to evolve. For example, there are debates on:

- whether the operational data store and the enterprise data warehouse truly represent different data environments or whether they are simply different names for the same type of data environment.
- when to use one database design versus another.

Continuing advancements in data storage, data processing, and software are another source of change. Because of the evolving nature of the subject, an instructor should expect to revise at least some of the definitions and concepts taught from one year to the next.

Another challenge is to incorporate some hands-on training so students obtain practical experience as well as a solid conceptual foundation. For realistic hands-on experience with data warehouses, students need access to the appropriate software and large amounts of data. One of the most important pieces of software in data warehousing is the DBMS. Although our students used Microsoft's Access for constructing their data warehouses, our department hopes in the future to switch to one of the leading DBMS products used in industry for data warehouse construction such as Oracle, SQL Server, or DB2. Once the DBMS is selected, the students may wish to explore other software products that assist in the design, data acquisition, administration, and data access and mining parts of the data warehouse. In addition to independent vendors, some of these tools are now being offered by the large DBMS vendors as part of their over all data warehouse software solution package. When budgets are tight, it is sometimes possible to obtain a free trial version of the software so that students can gain at least some exposure to the tools in this area. Table 5 lists vendors that offer free evaluation software and Table 6 lists vendors that offer web based or downloadable software demonstrations.

Table 5. Vendors that Offer Free Evaluation Software

Vendor	Web Site	Tool Name (Type)
Angoss Software	www.angoss.com	Knowledge Seeker (data mining tool) SmartWare 2000 (integrated product suite database/spreadsheet/wordprocessor)
Attar	www.attar.com	XPertRule (data mining tool)
Centura	www.centurasoft.com	SqlBase (relational data base)
Cognos	www.cognos.com	PowerPlay (OLAP analysis and reporting tool) Note: demonstration screens are available for other Cognos data mining products)
Computer Associates	www.cai.com	Ingres II (software development kit. Includes a relational database engine)
Hyperion	www.hyperion.com	Spider Man Web Application (web enabled DSS tool) Objects (family of OLAP aware components) Wired for OLAP (OLAP reporting, presentation tool)
IBM	www3.software.ibm.com/download	Numerous free trial versions are available.
Information Builders	www.ibi.com	WebFocus Developer (DSS Developer)
Lotus	www.lotus.com	Numerous free trial versions are available
Management Intelligenter Technologien	www.mitgmbh.de	Data Engine (data mining tool)
Megaputer Intelligence	www.megaputer.com	Numerous free trial versions are available (data mining tools).
Merant	www.merant.com	Data Direct (middleware, data access components)
Oracle	www.oracle.com/download	Numerous free trial versions are available.
Popkin Software	www.popkin.com	System Architect 2001 (data modeling tool)
Precise Software Solutions	www.precisesoft.com	Precise Enterprise Product Suite (performance tuning and monitoring tool)
RuleQuest	www.rulequest.com	See5 (data mining tool) Cubist (data mining tool)
Seagate	www.seagate.com	Seagate Info (access, analysis, and reporting tool) Seagate Analysis (OLAP, report design tool)
Silicon Graphics	www.sgi.com	MindSet (data mining tool)
SilverRun Technologies	www.silverrun.com	SR-BPM (Business Process Modeler) SR-ERX (Entity Relationship Expert) SR-RDM (Relational Data Modeler)
SPSS	www.spss.com/cool/index.htm	Numerous working copies of data analysis software are available.
Sybase	www.sybase.com	Power Designer (data modeling tool) and other Sybase products are available.
Viasoft	www.viasoft.com	Visual Process (knowledge base and repository tool)
Visible	www.visible.com	Visual Analyzer (data modeling tool) Visual Advantage (software engineering and repository tool)
WizSoft	www.wizsoft.com	WizWhy (data mining tool) WizRule (data quality tool)

Table 6. Vendors that Offer Web Based Or Downloadable Software Demonstrations

Vendor	Web Site	Tool Name (Type)
Business Objects	www.businessobiects.com	Business Objects (data access and reporting tool) WebIntelligence (web enabled DSS tool) Business Mining (data mining tool)
First Logic	www.firstlogic.com	i.d.centric (data quality tool) postalsoft (data quality tool)
Informix	www.informix.com	Informix product demos
SAS	www.sas.com	Enterprise Reporter (reporting tool) SAS/Graph (graphics tool) StatView (statistical analysis)
Speedware	www.speedware.com	Esperant (query and reporting tool)
Thinking Machines	www.think.com	Darwin (data mining tool)
VIT	www.vit.com	Enterprise Information Portal (information directory)

In terms of access to data, students can explore public online data warehouses. Besides being a source of data, these sites also provide students examples of data warehouses in action. Table 7 lists three examples of online data warehouses available to the public.

Table 7. Some Public Online Data Warehouses

nces.ed.gov/ipeds - IPEDS consists of institutional-level data that can be used to describe trends in higher education at the institutional, state and/or national levels.
www.census.gov - This is the web site of the U.S. Census Bureau. From the main page, students can gain access to data on the population, business, geography, and federal statistics.
www.nih.gov/health - This web site contains NIH health resources such as consumer health publications and databases, clinical trials, health hotlines, MEDLINE, and the NIH Information Index (a subject-word guide to diseases and conditions under investigation at NIH).

IX. STUDENT REACTION TO COURSE

At the end of each semester, students fill out a voluntary questionnaire to provide feedback as to how they felt about the course. Because the same instructor, textbooks, and syllabus were used throughout the year, the results from both the Fall and Spring semesters were pooled. Since the course was dual listed, the results are broken into two categories: undergraduate and graduate students. There were 30 students in the Fall semester class and 32 students in the Spring. By the end of the year, 28 senior MIS students and 34

MBA students had completed the course. The response rate for senior MIS students on the questionnaire was 68% while the response rate for MBA students was 76.5%.

In reviewing the results, a higher percentage of the graduate students (92.5%) found the textbooks and readings useful compared to the undergraduate students (68%). Younger students may have less experience reading industry trade books, which may account for the difference. Most undergraduate students (68%) felt the course load was about the same as other courses of equal credit compared to 46% of the graduate students. Interestingly, about a third of the graduate students (35%) felt the workload was more than their other courses. Roughly half of the undergraduate students (47%) and half of the graduate students (54%) felt the pace of the course was about right. Not surprisingly, more undergraduate students (42%) found the pace to be slightly fast compared to the graduate students (23%). Both the majorities of the undergraduate students (94.5%) and graduate students (88.5%) said they have learned something that they considered valuable. However, more graduate students (84.5%) said they would recommend the course to a friend compared to undergraduate students (68%). This suggests that this course may appeal more to the advanced business student.

In terms of miscellaneous comments, students said they liked those assignments and projects that were hands-on, practical in nature, and gave them a realistic feel for how they would construct a data warehouse in the corporate environment. They also enjoyed the speakers and real-life examples used during the lecture to illustrate the concepts being taught. Some students suggested that they would like to work on a large project that could be used as a unifying theme throughout the course. Other suggestions included switching from Access to Oracle, doing more group work, and expanding the data mining topics to a separate course.

Appendix B contains tables detailing the student responses.

X. SUMMARY

Data warehousing and mining represents new opportunities for course work in MIS. As an elective, data warehousing and mining gives students the opportunity to learn about data architecture on a corporate scale, to practice the steps involved with data warehouse construction, and to understand how organizations use data for business decision making. This course could be offered at either the undergraduate or graduate levels provided the students have a background in business, quantitative methods, and basic database skills.

This paper describes the type of content and activities that should be covered in the course. Some of the current challenges in offering the course include the lack of formal textbook support, the changing nature of the topic, and a lack of software and data for creating data warehouses on a realistic scale in the classroom. Despite these problems, this course offers students an important opportunity to explore further applications in databases and to appreciate the value of a cohesive data strategy for an organization.

Editor's Note: This paper was received on August 9, 1999. It was with the author 3 weeks for one revision. It was published on September 21, 1999

BIBLIOGRAPHY

Barquin, R. and H. Edelstein: (1997) *Building, Using & Managing the Data Warehouse*, Upper Saddle River, NJ: Prentice Hall.

Barquin, R. and H. Edelstein: (1997) *Planning and Designing the Data Warehouse*, Upper Saddle River, NJ: Prentice Hall.

Berry, M. and G. Linoff: (1997) *Data Mining Techniques For Marketing, Sales, and Customer Support*, New York, NY: John Wiley & Sons.

Bischoff, J. and T. Alexander: (1997) *Data Warehouse: Practical Advice from the Experts*, Upper Saddle Rive, NJ: Prentice Hall.

Date, C.: (1987) *An Introduction to Database Systems - Vol I (4th Edition)*, Reading, MA: Addison-Wesley.

Devlin, B.: (1997) *Data Warehouse from Architecture to Implementation*, Reading, MA: Addison-Wesley.

Gray, P. and H. J. Watson (1998) *Decision Support in the Data Warehouse*, Upper Saddle River, NJ: Prentice Hall PTR.

Groth, R.: (1998) *Data Mining: A Hands-On Approach for Business Professionals*, Upper Saddle River, NJ: Prentice Hall.

Inmon, W.: (1996) *Building the Data Warehouse (2nd Edition)*, New York, NY: John Wiley & Sons.

Inmon, W. and R. Hackathorn: (1994) *Using the Data Warehouse,* New York, NY: John Wiley & Son.

Inmon, W., C. Imhoff and G. Battas: (1996) *Building the Operational Data Store*, New York, NY: John Wiley & Sons.

Inmon, W., C. Imhoff, and R. Sousa: (1998) *Corporate Information Factory*, New York, NY: John Wiley & Sons.

Inmon, W., J. Welch and K. Glassey: (1997) *Managing the Data Warehouse*, New York, NY: John Wiley & Sons.

Kimball, R.: (1996) *The Data Warehouse Toolkit*, New York, NY: John Wiley & Sons.

Kimball, R., L. Reeves, M. Ross, and W. Thornthwaite: (1998) *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses*, New York, NY: John Wiley & Sons.

Mattison, R.: (1996) *Data Warehousing: Strategies, Technologies and Techniques*, New York, NY: McGraw-Hill.

Poe, V., P. Klauer. and S. Brobst: (1998) *Building a Data Warehouse for Decision Support*, Upper Saddle Rive, NJ: Prentice Hall .

Singh, H.: (1998) *Data Warehousing: Concepts, Technologies, Implementations, and Management*, Upper Saddle River, NJ: Prentice Hall.

Unica Technology, Inc.: (1997) *Solving Data Mining Problems Through Pattern Recognition*, Upper Saddle River, NJ: Prentice Hall.

APPENDIX A

A SAMPLING OF SCHOOLS THAT OFFER DATA WAREHOUSE AND DATA MINING COURSES

School	Course ID	Credit Hours	Book Used	Prerequisites
Cal Poly San Luis Obispo	BUS 491: Advanced Quantitative Methods and Control in Business	4	A, D, E, F	BUS 391: Management Information Systems
Claremont Graduate University	IS 304: Data Base Systems Planning	4	B	IS 302: Data Base Concepts
Drexel University	INFO 607: Applied Information and Database Technology	3	G, I	INFO 606: Database Management, INFO 620: Information Systems Analysis
Georgia Institute of Technology	CS 4440: Emerging Database Technologies	3	J	CS 4400: Introduction to Database Systems
Louisiana State University	ISDS 4110: Data Warehouse	3	C, G	ISDS 3110: Data Process Management
Raritan Valley Community College	CISY-287: Data Warehousing Fundamentals	3	H	CISY-285: Database Design & Development
UC San Diego	CSE 291: Topics in Database Systems: Information Integration, OLAP and Data Warehousing, Multimedia Information Systems	4	J	Essential knowledge about data structures, searching, and database systems is required.
University of Georgia	MGMT 7770: Data Warehouses and Mining	1.5	B	MGMT 7730: DSS
University of Houston	COSC 7397: Data Mining	3	J	COSC 6340: Data Management
University of South Florida	ISM 6930: Data Warehouse and Data Mining	3	G	Minimum 2 courses on relational DBMS

A - Elias and Awad: (1996) *Building Expert Systems: Principles, Procedures, and Applications*, St. Paul, MN: West Publishing Co.

B - Gray, P. and H. J. Watson (1998) *Decision Support in the Data Warehouse*, Upper Saddle River, NJ: Prentice Hall PTR.

C - Groth, R.: (1998) *Data Mining: A Hands-On Approach for Business Professionals*, Upper Saddle River, NJ: Prentice Hall.

D - Harjinder, G. and R. Prakash (1996): *The Official Guide to Data Warehousing*, Indianapolis, IN: Que Corporation.

E - Holsapple, C. and A. Whinston: (1996) *Decision Support Systems*, St. Paul, MN: West Publishing Co.

F - Inmon, W., C. Imhoff and G. Battas: (1996) *Building the Operational Data Store*, New York, NY: John Wiley & Sons.

G - Kimball, R.: (1996) *The Data Warehouse Toolkit*, New York, NY: John Wiley & Sons.

H - Kimball, R., L. Reeves, M. Ross, and W. Thornthwaite: (1998) *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses*, New York, NY: John Wiley & Sons.

I - Krakovsky, M. (1996): *Understanding the Oracle Server*, Upper Saddle River, NJ: Prentice Hall.

J - Selected papers from journals, conference proceedings, and books

APPENDIX B

STUDENT COURSE EVALUATIONS

Were the textbooks and readings useful?

	Strongly Agree	Agree	Disagree	Strongly Disagree	Other (Don't Know, Blank)
Undergrads	21%	47%	21%	10.5%	0%
Graduates	38.5%	54%	8%	0%	0%

Course workload as compared to other 3 credit courses?

	Much More	More	The Same	Less	Much Less
Undergrads	0%	16%	68%	16%	0%
Graduates	0%	35%	46%	19%	0%

Course pace?

	Too Slow	Slightly Slow	About Right	Slightly Fast	Too Fast
Undergrads	0%	5%	47%	42%	5%
Graduates	11.5%	11.5%	54%	23%	0%

Did you considered what you learned valuable?

	Strongly Agree	Agree	Disagree	Strongly Disagree	Other (Don't Know, Blank)
Undergrads	10.5%	84%	0%	5%	0%
Graduates	38.5%	50%	11.5%	0%	0%

Would you recommend this course to a friend?

	Strongly Agree	Agree	Disagree	Strongly Disagree	Other (Don't Know, Blank)
Undergrads	5%	63%	21%	5%	5%
Graduates	38.5%	46%	4%	8%	4%

ABOUT THE AUTHOR

E. M. Pierce joined the faculty of the MIS and Decision Sciences department at Indiana University of Pennsylvania in September 1995. She earned her Ph.D. from the Department of Statistics and Management Science in the School of Business at The University of Michigan in 1996. She also holds an M.S. in Computer Science from Iona College and a B.S. in Quantitative Business Analysis and Mathematics from The Pennsylvania State University. In addition to her academic career, she worked six years for IBM as an applications programmer developing and supporting internal business systems. Her research interests include decision support systems, data quality, data warehousing, and data mining.